Available online at https://www.iimk.ac.in/research/markconf20/

INDIAN INSTITUTE OF MANAGEMENT KOZHIKODE

04th International Conference on Marketing, Technology & Society 2020

Sentimental Analysis using Text Mining of Selecting Location of Flat Buyers in Chennai

Dr.S.Ajitha.[1]
Murali K[2]
Joel Andrew M[3]

Abstract

The prior researches suggest about the price fixing of the property which predicts and calculate price of property by sentiment analysis using text mining. The purpose of this paper to explain about the analysis of people buying flat in Chennai by sentiment analysis using text mining is known to be ("SATLFB"). The study focus on the location preference of flat buyers in Chennai. Since there are many parameters to select a location for buying flat in Chennai which is life time savings of a person in India. There are many parameter such as age, price, locations based on temple, entertainment such well-planned, future development, employments etc. we are going to do analysis from top Six data of these parameter which a buyer giving preference to buy flat in Chennai .The sentiment analysis uses reviews from magicbricks.com. The data collected from Jan 2018 –Jan 2019 for the above parameter Results show SATLFB how useful to analysis the flat buyers mind and preference of location in Chennai by the above parameter .The paper useful to know the people mentally about selection of location in Chennai. The reliability of parameter which buyers use to select the location to buy the flat, if they don't have exact information of parameter sometimes the assumption may go wrong which may lead to buying flat in wrong location. The implication of this paper can lead people to buy flat in location actually they want. This may change the buying method of people since major data collected here are accurate. The system provides information to sort out or to identify the location where the buyers to buy flat in Chennai according to their needs.

Keywords: sentiment analysis, text mining, property buyers.

1. Introduction

The flat buyers in Chennai have many criteria to select location where to buy the flat. If they don't have idea about the location it will lead to wrong investment for life time. There are many factors such as devotional place, well planned, future development, employments, school and transport. These are above six major factors in which people use to sort out the location where they want to buy house.

Sentimental analysis is done using text mining for selection of location for buying flat. The data used for the analysis is taken from magicbricks.com .The data from Magic bricks is retired and sentimental analysis is done.

2. Literature Review

This paper discusses the potential applications and methodological challenges of using Latent Semantic Analysis (LSA) to research unstructured data in land research. LSA is a statistical tool that permits for the analysis of huge bodies of textual data by identifying specific topics or themes within the data. An in-depth discussion of the analytic process as a series of specific steps is provided, followed by three applications that illustrate the methodological decisions made by the analyst when LSA is performed. In this study, so as to research user needs for land, they focus on "Mansion Community" which is a Japanese bulletin board system (hereinafter referred to as BBS) about Japanese real estate. In the study, extraction of keywords is performed supported calculation of the entropy value of every word,

[1] *Assistant professor, ITM Business School Email: ajithas@itm.edu*
[2] *ITM Business School Email:muralivishnu086@gmail.com*
[3] *ITM Business School Email: mjoelandrew@gmail.com*

Peer-review under responsibility of the 04th ICMTS 2020
ISBN : 978-93-5419-748-2

and that we used them as features during a machine learning classifier to analyze 6 million posts at "Mansion Community". They explore whether social media can provide a window into community land — foreclosure rates and price changes — beyond that of traditional economic and demographic variables. The paper investigates the role of fundamentals and investor sentiment in commercial real estate valuation. In land markets, heterogeneous properties trade illiquid, highly segmented and informationally inefficient local markets. Moreover, the lack to short sell private land restricts the power of sophisticated traders to enter the market and eliminate mispricing. These characteristics would appear to render private land markets highly vulnerable to sentiment-induced mispricing. Using error correction models to carefully model potential lags within the adjustment process. They also show how our model are often wont to highlight keywords that affect the worth positively or negatively. The analysis shows that our model (which exploits textual features) achieves significantly lower root mean squared error across the various data sets and against sort of regression models. Analyze the important estate transaction data, and built prediction models for the important estate price using data processing algorithms, and validate the models. They built two models using two algorithms - decision trees and neural networks - and compared their performances. The process of mining real-estate listings based on decision systems over ontological graphs. Such decision systems are proposed to affect data within the sort of concepts linked by different semantic relations. A special attention is concentrated on pre-processing steps transforming advertisements within the textual form into decision systems (decision tables) defined over ontological graphs.

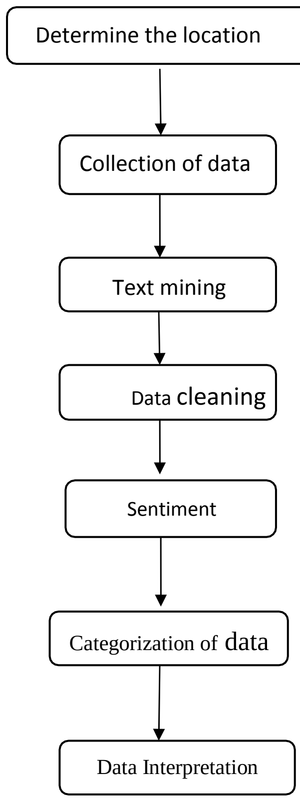
3. Methodology

The research is done by using descriptive qualitative method. The stages of this research is given below as a drawing.

Text mining is used to gather information and categories theses information, group them and perform sentimental analysis by using these data and prepare summary of the result. In this paper text mining using reviews from Magicbricks.com. Data cleaning is done to perform what to be done in this case study, in location selection by flat buyers. In the location selection process we can find how magicbricks.com reviews help to the location by the buyers.

Sentimental analysis is mix of statistics machine language and natural language processing (NLP) to extract information from text files. It used by buyers by both way negatively and positively such cases some may avoid buying flats there. But main purpose of this paper is to make easy to buyers to short list the location where they need to buy the flats.

.



3.1 Data Retrieval

Data here is nothing but reviews from magicbricks.com which is done by python. By using python reviews of each location is retired from website initially. We get data in HTML format then we are converting them to required form.

3.2 Text Vectorization

Data retrieved from magic bricks is processed by text vectorization. Here we use count vectorization method to analyze the data from magic bricks web site. It is nothing but separating the reviews from web site and separating the reviews into vectors and count them for the analyze of reviews or data. By the count the particular word we can do analyze the people’s reviews about the particular location by different aspects such as companies, parks, schools and infrastructure in the location.

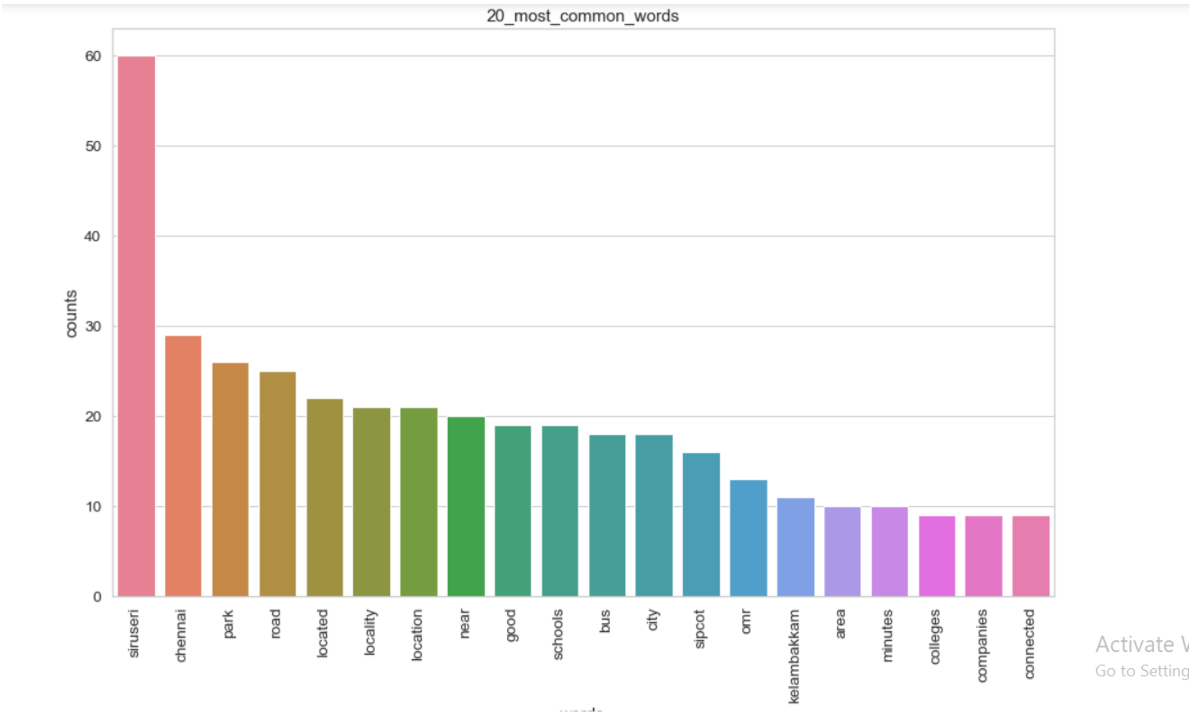
By using the count vectorization we can analyze the count of words which occurred as many as will come in first which shows that it was the reviews of people given majorly.

4. Findings

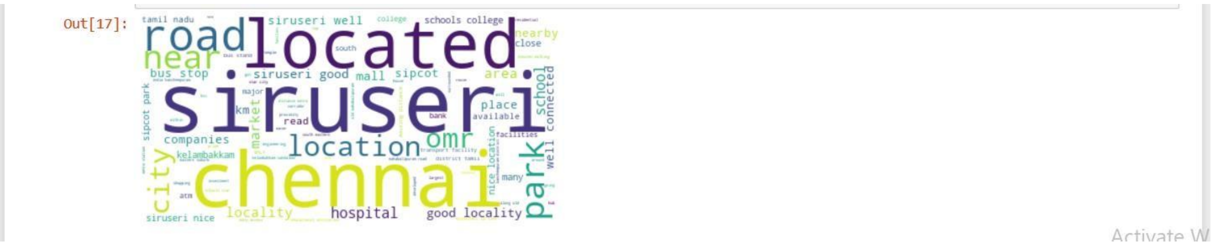
We did this procedure one location namely siruseri. We got totally 49 reviews about this location by using python.

Then the sentimental analyzed by using count vectorization the review is then segregated to vector and count is used to each term and by using this count ,count vectorization and word cloud. By using the below coding we can achieve what we need.

```
import pandas as pd import os os.chdir('..')
siruseri=pd.read_csv(r'./Downloads/siruseri1.csv') siruseri.head()
siruseri=siruseri.drop(columns=['web-scraper-order','web-scraper-start-url'],axis=1) import re
siruseri = siruseri['review'].map(lambda x: re.sub('[,\.!?]','',x)) siruseri=siruseri.map(lambda x: x.lower())
from wordcloud.wordcloud import WordCloud import matplotlib long_string=','.join(list(siruseri.values))
wordcloud=WordCloud(background_color="white",max_words=1000,contour_width=3,cont
our_color='steelblue')
wordcloud.generate(long_string) wordcloud.to_image()
from sklearn.feature_extraction.text import CountVectorizer import numpy as np
import matplotlib. pyplot as plt import seaborn as sns
sns. set_style( 'whitegrid')
%matplotlib inline # Helper function
def plot_20_most_common_words(count_data, count_vectorizer) : import matplotlib.pyplot as plt
words=count_vectorizer . get_feature_names() total_counts=np.zeros (len (words))
for t in count_data: total_counts+=t.toarray() [0] count_dict=(zip(words, total_counts))
count_dict = sorted (count_dict, key=lambda x:x[1], reverse=True) [0: 20] words=[w[0] for w in
count_dict]
counts=[w[1] for w in count_dict] x_pos=np.arange(len(words)) plt.figure(2, figsize=(15, 15/1.6180))
plt.subplot (title= '20_most_common_words')
sns.set_context("notebook", font_scale=1.25,rc={"lines . linewidth " :2.5}) sns .barplot(x_pos, counts,
palette= 'husl')
plt .xticks(x_pos, words, rotation=90) plt .xlabel( 'words' )
plt .ylabel( 'counts') plt .show()
# Initialise the count vectorizer with English stop words count_vectorizer =
CountVectorizer(stop_words='english') # Fit and transform the processed titles
count_data=count_vectorizer.fit_transform (siruseri)
# Visualise the 10 most common words plot_20_most_common_words(count_data, count_vectorizer)
```



Count vectorization



Word cloud

5. Conclusion

By using above results a buyer who is planning to buy a flat in Chennai can come to a conclusion regarding his need and factors regarding the choosing location of flat buying in Chennai

References

1. Sherief Abdallaha (2015) ‘Using Text Mining To Analyze Real Estate Classifieds’
2. Victor Gan, Vaishali Agarwal, Ben Kim(2015) ‘Data Mining analysis and predictions of real estate prices’
3. Krzysztof Pancerz and Olga Mich , ‘Mining Real-Estate Listings Based on Decision Systems over Ontological Graphs’
4. Nicholas Evangelopoulos, Kimberly Winson-Geideman, Triss Ashton, Stephen Roulac (2015) ‘Latent semantic analysis and real estate research: Methods and applications’.
5. Hiroki Horino, Hirofumi Nonaka, Elisa Claire Alemán Carreón, Toru Hiraoka ‘Development of an Entropy-Based Feature Selection Method and Analysis of Online Reviews on Real Estate’.
6. MohammadzamanZamani and HansenAndrewSchwartz ‘Using Twitter Language to Predict the Real Estate Market’.
7. Jim Clayton & David C. Ling & Andy Naranjo (2008) ‘Commercial Real Estate Valuation: Fundamentals Versus Investor Sentiment’.
8. Dini Turipanam Alamanda, Abdullah Ramdhani, Ikeu Kania, Wati Susilawati, Egi Septian Hadi ‘Sentiment Analysis Using Text Mining of Indonesia Tourism Reviews via Social Media’